

Title: "Interactive Classification Technology"

Type: Summary Report, submitted March 27, 2001

Principal Investigator: Cary deBessonnet

Period Covered: 1/1/99 - 12/31/00

Institution: Southern University Law Center
P.O. Box 9294
Baton Rouge, LA 70813

Grant No: NAG 2-1273

1.1 Invention Statement

No subject inventions resulted from the research.

1.2 Goals, Accomplishments and Findings

The investigators upgraded a knowledge representation language called SL (Symbolic Language) and an automated reasoning system called SMS (Symbolic Manipulation System) to enable the more effective use of the technologies in automated reasoning and interactive classification systems. The overall goals of the project were:

- 1) the enhancement of the representation language SL to accommodate a wider range of meaning;
- 2) the development of a default inference scheme to operate over SL notation as it is encoded; and
- 3) the development of an interpreter for SL that would handle representations of some basic cognitive acts and perspectives.

Much of the work focused on inferencing and representation issues, including the development of high-level classification functions and conceptual models as well as the specification of a detailed scheme of defaults designed to operate over SL notation as it is encoded. The technology was developed in the context of an SMS environment, that is, a universe of knowledge bases (KBs) built by the SMS interpreter. The KBs are built from epistemic elements carried by SL expressions, usually in the form of representations of objects, states and events distinguished from one another based on semantics and syntax. The KBs function as the "reality" of the system, and thus the system can be said to employ knowledge-base semantics (Hirst, 1987). The "reality" so constructed serves as the world model into which mappings from SL sentences can be made when evaluating SL expressions from a semantic perspective. This "reality" is conceived of as a complex of objects and relations distributed among ontologically diverse realms (cf. worlds) over which inferencing is defined. Ontological diversity is captured by employing a system of assignment by which values are assigned to objects to indicate their ontological status. Each sentential object of "reality" is *signed*, that is, each object carries one or more values that describe its status or presence (cf. truth) within the system. Values so assigned in this many-valued context can amount to qualifications imposed on the participation of objects in the "reality" constructed by the system. Even so, the reasoning is distinct from fuzzy reasoning (e.g. see Yager *et al* for collective discussion of Zadeh's works) because objects remain "crisp" (not fuzzy) in the realms in which they appear (deBessonnet, 1991; 1995). This is not to say that the technology is incompatible with a fuzzy approach, or with a probabilistic approach for that matter, only that fuzzy and probabilistic distributions would have to be made available and incorporated into the system prior to their being used in the usual way.

In the approach developed, interpretive capabilities constitute an integral part of artificial cognition. The interpretation of a sentence of SL is partly a function of the interpretation of its components. Interpretation is performed by functions that operate on: atoms; complex phrases; sentences; and knowledge bases. With respect to the interpretation of knowledge bases, the first level of interpretation results in a top level KB consisting of well-formed sentences of SL. Thereafter the components of the sentences are individuated and cast in to an equivalent KB that is optimized for storage and retrieval. This KB is accessed in query/response mode to respond to queries posed by the user.

Since much of what takes place within ordinary human conversation can hardly be said to be analytic or deductive in nature, the system was designed to handle nonstandard, even tainted, inferencing. In

SMS, tainted inferencing is referred to as “penumbral inferencing” and is stipulatively defined as inferencing that is neither deductive nor probabilistic. Although penumbral inferences are nonstandard, they are quite useful and are seemingly indispensable for the successful operation of technology of this kind. Sentential objects of penumbral regions are tainted in recognizable ways. Tainted inferences are classified according to type, and numerous sources of imprecision are recognized in the system, among them: 1) disjunctive information (e.g. “John or Mary,” but which one?); 2) indefinite existential import (e.g. “unicorns,” but do these objects exist?); 3) indefinite realm (e.g. “bigfoot,” but is the object a member of the fictitious world?); and indefinite relations between sets (e.g. “women saw men,” but which one or more women saw which one or more men?).

The investigators developed a proof methodology that proceeds in bottom-up fashion to map components of queries into SMS KBs in the first round of evaluation. The system is based on a theory of proof developed by the principal investigator by which query components are mapped independently into the knowledge bases (see deBessonnet, 1991, section 11.5). This enables the system to keep track of query satisfactions at the atomic level, which can then be appreciated at the sentential level. In subsequent rounds of evaluation, surviving component satisfactions are related and tested with tractability in view. The system attempts to satisfy queries from multiple perspectives and searches penumbral regions in the process. Search paths are taken to be lines of reasoning that lead to the presented query, which is taken as a proposition to be proved. Since a line of reasoning might involve steps that are not deductive, the system keeps track of those instances so that it can appropriately qualify its responses.

The technology was developed to process factual information (descriptions) with a view towards using it generally in rule-based systems, particularly in those designed to function as “cogitating” mechanisms, such as robots. Most of the research during the reporting period focused on goals (1) and (2). The encoding and processing of perspectives were treated as part of the specification of syntax and rules of inference; hence, issues presented by goal (3) were addressed as part of the effort to achieve the first two goals. In SMS, perspectives include both individual and global points of view. The system interpreter maintains a global point of view, for example. The functions and models required to achieve goal (3), including the treatment of perspectives, depend on the availability of a notation that is capable of accommodating both models and perspectives. Accordingly, much of the research effort was spent on the enhancement of SL in these areas. The syntax rules of SL, for instance, must strike an appropriate balance between providing information by means of explicit encoding as opposed to inclusion by default. This consideration is particularly important since, to achieve the goals, SMS must allow for multiple perspectives to be encoded simultaneously. It was found that allowing the scale to dip too heavily in the direction of explicit encoding results in the notation becoming too tedious for practical use.

The investigators developed a scheme by which SMS recognizes objects when they become individuated, that is, distinguished by a unique marker. The objects are classified using multiple bases, such as whether they are:

- 1) divisible or indivisible;
- 2) commutable or not;
- 3) associative or not;
- 4) transportable or not.
- 5) an abstract concept or an instance of a concept;
- 6) a single entity or a group;
- 7) a fully present object or a penumbral object; and
- 8) a class or not.

Each of these bases is defined for particular purposes, and each affects inferencing in specified ways. The classification of an object as “divisible” or “indivisible,” for example, determines whether it may be divided without affecting its ontological status (e.g. its existence in a particular realm). Disjunctive objects, such as a sequence of the form $\langle ?a \text{ or } ?b \text{ or } ?c \text{ or } \dots ?n \rangle$ (where “?” flags variables) are indivisible within the realm of OW (for “Ordinary World”). Such an object cannot be divided without the results being transported to another realm. This object, although indivisible in the realm of OW, is both commutable and associative within that realm. The effects produced are obvious. Given the presence of the assertion $\langle \langle \text{john or mary or jim} \rangle \text{ is tall} \rangle$ in the realm of OW, the following inferences would be allowed in OW without the results being transported to another realm:

- a) the commutative inference $\langle \langle \text{mary or jim} \rangle \text{ or john} \rangle \text{ is tall} \rangle$; and
- b) the associative inference $\langle \langle \langle \text{john or mary} \rangle \text{ or jim} \rangle \text{ is tall} \rangle$.

On the other hand, the inference '<john is tall>' (which would result from division) would not be allowed and, as such, would be transported to another realm, namely, DW (for "Disjunctive World"), a realm that houses objects tainted by the fact that their assertion into "reality" was based on disjunction rather than direct assertion. An assertion such as '<john caused <mary is happy>>' would be neither commutable nor associative in OW.

After sufficient progress was made in the improvement of SL, new features were added to SMS to enable the system to handle extended SL notation. Special functions and models were developed for use in interpretation and inferencing over a variety of SL sentence types, including those that employ special modal syntax or operators. Since SL is being designed as a general representation language capable of handling a wide range of objects, including processes and events, the investigators developed techniques by which SMS can recognize and incorporate ontological distinctions into its knowledge bases. Considerable efforts were directed during the research term towards improving the inferencing and retrieval capabilities of the system to enable it to operate in a conversational mode in which the user could pose questions and have the system generate responses based on the contents of its knowledge bases. The results of that effort brought about important progress. The research conducted over the reporting period convincingly confirmed suspicions that for SMS to reach its potential, it must employ a many-valued scheme of inference and must be able to map out its lines of reasoning for the user.

2.0 Summary of Accomplishments

Accomplishments during the reporting period include:

- 1) the specification of a more flexible syntax and semantic theory for the language;
- 2) the development of a syntax checker to ease encoding tasks; and
- 3) the development of a broader range of syntactical functions and models for use in inferencing over a variety of SL sentence types, including those that employ special modal syntax or operators (the modal operators were developed for both sentential objects as well as their components).
- 4) the development of tractable inferencing and retrieval operations.

The investigators were also successful in developing techniques by which the system is in a position to explain its responses. This was found to be a valuable feature because the system sometimes makes quite fine distinctions in inferencing that perhaps would be overlooked by a human being.

BIBLIOGRAPHY

- deBessonnet, C. G., (1995a), "Towards a sentential reality for the android," in K. Ford, C. Glymour and P. Hayes (eds.), *Android Epistemology*, AAAI/MIT Press.
- deBessonnet, C. G. (1991). *A Many-Valued Approach to Deduction and Reasoning for Artificial Intelligence*. Norwell, MA: Kluwer Academic Publishers.
- Hayes, P. H. (1994). What the frame problem is and isn't. In Z. Pylyshyn, Ed. *The Robot's Dilemma: the Frame Problem in Artificial Intelligence*. Ablex Publishing Co.
- Hayes, P. H. (1987). Aristotelian and Platonic views of knowledge representation. *Lecture Notes in Artificial Intelligence*, 835, Springer, pp. 1-11.
- Hirst, G. (1987), *Semantic Interpretation and the Resolution of Ambiguity*, Cambridge University Press, Cambridge.
- Hobbs, J. R. (1985), "Ontological promiscuity," *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, Chicago, Illinois, 1985, pp. 61-69.
- Pearl, J (1988), *Networks of Belief: Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, Los Altos, California.
- Yager, R., Ovchinnikov, S., Tong, R., and Nguyen, H. (eds.) (1987), *Fuzzy Sets and Applications: Selected Papers by L. A. Zadeh*, John Wiley & Sons, New York.